

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
18 April 2002 (18.04.2002)

PCT

(10) International Publication Number  
**WO 02/31613 A2**

(51) International Patent Classification<sup>7</sup>:

G06F

(74) Agent: G. E. EHRLICH (1995) LTD.; 28 Bezael Street,  
52521 Ramat Gan (IL).

(21) International Application Number: PCT/IL01/00937

(22) International Filing Date: 10 October 2001 (10.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/689,884 13 October 2000 (13.10.2000) US

(71) Applicant (for all designated States except US): INSYST  
LTD. [IL/IL]; Har Hotzvim, 7 HaMarpe Street, 91231  
Jerusalem (IL).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI,  
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU,  
ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG).

(72) Inventors; and

(75) Inventors/Applicants (for US only): GOLDMAN,  
Arnold, J. [IL/IL]; 18 HaPalmach Street, 92542 Jerusalem  
(IL). HARTMAN, Jehuda [IL/IL]; 18 HaTzanchanim  
Street, 76212 Rehovot (IL). FISHER, Joseph [IL/IL]; 3  
Mishol HeHadas Street, 97277 Jerusalem (IL). SAREL,  
Shlomo [IL/IL]; 151 Ma'ale MichMash, 90634 M.P. Yash  
(IL).

Published:

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

WO 02/31613 A2

(54) Title: SYSTEM AND METHOD FOR MONITORING PROCESS QUALITY CONTROL

(57) Abstract: A method of modeling a monitorable stage in a process is provided. The method including the steps of: (a) mea-  
suring at least one input value of a parameter of the monitorable stage of the process; (a) measuring at least one output value of  
the parameter of the monitorable stage of the process; and (c) utilizing the at least one input value and the at least one output value  
for constructing a process output empirical model for uncovering a functional relationship between the at least one input value and  
at least one output value, the step of constructing the process output empirical modeler being effected by: (i) dividing at least one  
interval of the parameter into a plurality of sub intervals, such that each of the at least one interval is divided into at least two of the  
sub intervals; (ii) classifying the at least one output value according to the plurality of sub intervals, thereby presenting the at least  
one output value as a plurality of discrete variables defining the at least one output value; and (iii) using the plurality of discrete  
variables defining the at least one output value for defining the functional relationship between the at least one input value and the at  
least one output value, thereby modeling the monitorable stage of the process.

5

SYSTEM AND METHOD FOR MONITORING PROCESS  
QUALITY CONTROL

FIELD OF THE INVENTION

10

The present invention relates to a system and method for automatic learning and rule induction from data. It could be applied to any situation where a cause and effect relationship between a plurality of input parameters and an output parameter, and historical data of the said input and output parameters is available. When applied to a process, the present invention relates to a system and method for monitoring and optimizing process quality control and, more particularly but not exclusively, to a system and method which employs an algorithm to provide a model useful for accurate and sensitive monitoring of a process, which enables detection of parameter(s) deviation even at early stages of a process.

15  
20

BACKGROUND OF THE INVENTION

In many areas and situations a cause and effect relationship between a plurality of input parameters and an output value

exists. The present invention relates to a system and method for automatic learning and rule induction from data. More specifically, the present invention is a system and method to uncover the multivariate functional relationship between the input and output parameters. This function constitutes an empirical model of the relationship. It could be applied to any situation where historical data of input and output parameters is available. One of the areas that the present invention is applied is Process Quality Control. Traditionally, quality control of simple processes involves the classification of end products. In more complicated processes, which utilize numerous process stages, some quality control is affected in intermediate stages, involving the classification of intermediate products.

For example, in a chemical process, which includes numerous stages, inspection samples are typically drawn at random at various stages of the production line and inspected for being within predefined control limits.

A quality control methodology which is indicative of the quality of end products, is at times unacceptable for some processes since it cannot detect variabilities in intermediates produced.

Some processes, such as those employed by the semiconductor industry, utilize statistical process control (SPC), which uses control charts to analyze each major process stage and generate a predictable distribution chart for measured parameters (outputs) at each stage. A measured parameter which

deviates from its distribution chart by more than, for example, three standard deviations is taken as indicative of process problems.

Although such quality control far supersedes that effected by sample inspection, it still suffers from several inherent limitations. The main  
5 reason is that the traditional SPC monitors an output with respect to the entire statistical distribution of this output. Each input combination defines a distribution of the related output, thus the overall distribution consists of many (sub) distributions.

By monitoring outputs with their own specific distribution we achieve a  
10 much higher degree of accuracy. For example, the distribution charts of process outputs at various stages cannot detect undesirable combinations of input variables (e.g. such in which the unfavorable effect of the inputs on the monitored process output are mutually compensated), as long as the process outputs are within specifications. As a result, such quality control methodology  
15 cannot be utilized for early detection of variability in a process, nor can it be utilized to detect and point out deviations in individual variables, which may be important for understanding process related problems.

There is thus a widely recognized need for, and it would be highly advantageous to have, a system and method for process quality control devoid  
20 of the above limitations.

SUMMARY OF THE INVENTION

According to one aspect of the present invention there is provided a method of modeling a monitorable stage in a process, the method comprising the steps of: (a) measuring at least one input value of a parameter of the monitorable stage of the process; (b) measuring at least one output value of the parameter of the monitorable stage of the process; and (c) utilizing the at least one input value and the at least one output value for constructing a process output empirical model for uncovering a functional relationship between the at least one input value and at least one output value, the step of constructing the process output empirical modeler being effected by: (i) dividing at least one interval of the parameter into a plurality of sub intervals, such that each of the at least one interval is divided into at least two of the sub intervals; (ii) classifying the at least one output value according to the plurality of sub intervals, thereby presenting the at least one output value as a plurality of discrete variables defining the at least one output value; and (iii) using the plurality of discrete variables defining the at least one output value for defining the functional relationship between the at least one input value and the at least one output value, thereby modeling the monitorable stage of the process.

According to another aspect of the present invention there is provided a method of assessing the quality of a monitorable stage of a process, the method comprising the steps of: (a) constructing a process output empirical model for uncovering a functional relationship between an

input value and an output value of a parameter of the monitorable stage of the process, the step of constructing a process output empirical model being effected by: (i) dividing at least one interval of the parameter into a plurality of sub intervals, such that each of the at least one interval is divided into at least two of the sub intervals;

(ii) classifying at least one output value according to the plurality of sub intervals, thereby presenting the at least one output value as a plurality of discrete variables defining the at least one output value; and (iii) using the plurality of discrete variables defining the at least one output value for defining a functional relationship between at least one input value and at least one output value, thereby modeling the monitorable stage in the process; (b) applying the process output empirical model to a measured input value of the monitorable stage so as to predict a distribution of the output value of the monitorable stage; and (c) comparing a measured output value of the monitorable stage to the distribution of the output value of the monitorable stage predicted in step (b) to thereby assess the quality of the monitorable stage of the process.

According to yet another aspect of the present invention there is provided a system for assessing the quality of a process, the system comprising a data processing unit being for: (a) receiving a measured input value of a parameter of a monitorable stage of the process; (b) predicting a distribution of an output value of the parameter of the monitorable stage of the process according to the measured input value, the step of predicting

being effected by a process output empirical model being executed by the data processing unit, the process output empirical model being generated by: (i) dividing at least one interval of the parameter into a plurality of sub intervals, such that each of the at least one interval is divided into at least two of the sub intervals; (ii) classifying at least one output value of the parameter according to the plurality of sub intervals, thereby presenting the at least one output value as a plurality of discrete variables defining the at least one output value; and (iii) using the plurality of discrete variables defining the at least one output value for defining the functional relationship between the at least one input value and at least one output value; and (c) comparing a measured output value of the parameter to the distribution of the output value of the parameter predicted in step (b), to thereby assess the quality of the monitorable stage of the process.

According to further features in preferred embodiments of the invention described below, each sub interval of the at least two sub intervals encompasses a non-overlapping subset of output values.

According to still further features in the described preferred embodiments the functional relationship is defined via a discrete function.

According to still further features in the described preferred embodiments the step of constructing the process output empirical modeler further includes the step of: (iv) statistically testing the discrete function for the goodness of the statistical result.

According to still further features in the described preferred embodiments the process is selected from the group consisting of a medical diagnostic process, a wafer production process and a trade order execution process.

5           According to still further features in the described preferred embodiments the monitorable stage of the process is a wafer chemical mechanical polishing stage of a wafer production process.

          According to still further features in the described preferred embodiments the system further comprising at least one sensor being in  
10           communication with the data processing unit, the at least one sensor being for collecting data from the monitorable stage of the process, the data including the at least one input value and the at least one output value of the parameter.

          According to yet an additional aspect of the present invention  
15           there is provided a method of assessing the quality of a monitorable stage of a process, the method comprising the steps of: (a) processing at least one output value of a parameter of the monitorable stage of the process so as to generate discrete variables representing the at least one output value; (b) defining a function for associating the discrete variables and at least one  
20           input value of the parameter of the monitorable stage of the process; (c) applying the function to a measured input value of the monitorable stage so as to predict a distribution of the output value of the monitorable stage; and (d) comparing a measured output value of the monitorable stage to the



distribution of the output value of the monitorable stage predicted in step (c) to thereby assess the quality of the monitorable stage of the process.

According to still an additional aspect of the present invention there is provided a system for assessing the quality of a monitorable stage of a process, the system comprising a data processing unit being for: (a) processing at least one output value of a parameter of the monitorable stage of the process so as to generate discrete variables representing the at least one output value; (b) defining a function for associating the discrete variables and at least one input value of the parameter of the monitorable stage of the process; (c) applying the function to a measured input value of the monitorable stage so as to predict a distribution of the output value of the monitorable stage; and (d) comparing a measured output value of the monitorable stage to the distribution of the output value of the monitorable stage predicted in step (c) to thereby assess the quality of the monitorable stage of the process.

According to still further features in the described preferred embodiments the function is defined via non-parametric statistics.

According to still further features in the described preferred embodiments the function is a discrete function.

According to still further features in the described preferred embodiments the discrete variables are generated by dividing at least one interval of the parameter into a plurality of sub intervals and classifying the at least one output value according to the plurality of sub intervals.

According to still further features in the described preferred embodiments the system further comprising at least one sensor being in communication with the data processing unit, the at least one sensor being for collecting data from the monitorable stage of the process, the data  
5 including the at least one input value and the at least one output value of the parameter.

Embodiments of the invention address the shortcomings of the presently known configurations by providing a system and method for  
10 assessing the quality of at least one monitorable stage of a process thus enabling to optimize the process in a model which is useful for accurate and sensitive monitoring of the process. The model preferably enables detection of parameter(s) deviation even at early stages of the process

#### 15 BRIEF DESCRIPTION OF THE DRAWINGS

The invention is herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of the preferred  
20 embodiments of the present invention only, and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of the invention. In this regard, no attempt is made to show structural details of the invention in

more detail than is necessary for a fundamental understanding of the invention, the description taken with the drawings making apparent to those skilled in the art how the several forms of the invention may be embodied in practice.

5 In the drawings:

FIG. 1 is a generalized block diagram showing a first embodiment of the present invention configured in a learning mode,

FIG. 2 is a generalized flow diagram of the learning state of the embodiment of Fig. 1,

10 FIG. 3 is a generalized block diagram of a process control state of the embodiment of Fig. 1,

FIG. 4 is a generalized flow diagram of the process control state of Fig. 3,

15 FIG. 5 is a generalized flow diagram showing how a model built using the learning mode of Fig. 1, can be used to obtain an understanding of a process,

FIG. 6 represents a cause and effect functional relationship having six inputs (process variables), each variable interval is divided to three sub intervals (A, B and C) and graph depicting for various input combinations the process output distribution according to the teachings of  
20 the present invention,

FIG. 7 illustrates the discretization of the four input streams and the assignment of different output distributions to each input (vector) combination,

FIG. 8 is an example of a feedback control loop in the semiconductor industry implemented by the present invention,

FIG. 9 shows a table of raw data collected during a chemical mechanical polishing (CMP) stage of wafer production,

FIG. 10 shows input vectors construction in the implementation of a process output empirical modeler (POEM) to the process shown in FIG. 8,

FIG. 11 shows a look-up table generated by the algorithm of the present invention, which is useful for predicting a distribution of an output value according to a measured input value,

FIG. 12 is a window of a graphical interface during the computerized monitoring and control of the process shown in FIG. 8,

FIG. 13 illustrates an improvement achieved by applying the process output empirical modeler (POEM) to the CMP machine,

FIG. 14 is a medical example of uncovering the quantitative relationship of the likelihood of a pathology as function of four tests and the patient's history from historical data,

FIG. 15 is an example of a cause and effect medical relationship with seven input variables and two outputs, and

Figs. 16a - c are simplified drawings illustrating a further embodiment of the present invention.

### DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is of a system and method, which can be utilized to optimize at least one stage of a process. Specifically, the present invention can be used to generate a model for functionally relating input and output values of a parameter of the at least one stage in a process so as to enable prediction of a distribution of an output value based on an input value measured from the process.

The principles and operation of the present invention may be better understood with reference to the drawings and accompanying descriptions.

Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details of construction and the arrangement of the components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments or of being practiced or carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein is for the purpose of description and should not be regarded as limiting.

It will be appreciated that any process stage, which includes measurable inputs and outputs, can be modeled and assessed for quality and thus optimized utilizing the process output empirical modeler of the present invention. Examples of such processes include but are not limited to,

medical diagnostic processes, such as the diagnosis of pathologies according to blood tests, wafer production processes, such as the chemical polishing stage of wafer production, or trade order execution processes. The application of the process output empirical modeler to such processes is described in detail in the Examples section which follows.

Additional objects, advantages, and novel features of the present invention will become apparent to one ordinarily skilled in the art upon examination of the following examples, which are not intended to be limiting. Additionally, each of the various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below finds experimental support in the following examples.

Reference is now made to Fig. 1, which shows a system according to a first embodiment of the present invention configured in a learning mode. Generally a system according to embodiments of the invention has a learning mode during which it collects and arranges input and output data of a process in order to develop a model and an operating mode during which it monitors a process according to the model developed during the learning mode.

The model takes a series of inputs and at least one output, and follows the process for a statistically significant period of time so that an empirical relationship can be built up between different values at the inputs and measured output values. In Fig. 1, a series of input parameters  $I_1 \dots I_n$  are each assigned expected ranges in a parameter definition unit 10. The

expected ranges are discretized into subranges by a range divider 12 and then a series of vectors is formed of each possible combination of subranges in a vector former 14. For example if there are three inputs and each input is divided into three subranges then 27 vectors are formed.

5           Each one of the 27 vectors thus covers a certain part of the input space and will correspond to a certain portion of the output space.

Now the process is allowed to start and measurements are made of actual input and corresponding output values in a measurement input unit 16. A result categorizer 18 then takes each y output measurement and looks  
10   at the corresponding inputs that gave rise thereto. Each one of the corresponding inputs may be associated with a subrange as defined by the range divider and thus each input y value may be associated with one of the vectors.

Measurement is continued until it is felt that a statistically  
15   significant sample of results is built up. This may be after many measurements. In particular it is preferable that there should be enough y results associated with each vector to give a meaningful statistical distribution per vector. Thus each vector should have a large number of y results associated therewith. There should be enough y results associated  
20   with each vector to give a meaningful statistical distribution per vector. The statistical distribution per vector may be studied in a statistical analysis unit 20. Statistical analysis preferably includes using some kind of score to indicate the goodness of the statistical results.

Following statistical analysis of each vector annealing of the vectors is carried out, in an annealing unit 22 by changing the boundaries between subranges. The y results are then reassigned to the annealed vectors by the results categorizer 18 and the statistical analysis is repeated, a  
5 new score being calculated. If the score is better than previously the new boundaries are accepted. The loop is repeated until a condition is fulfilled which indicates that the best possible result has been found. Several possible types of condition will suggest themselves to the skilled person. Preferably the condition chosen will not allow calculation to stop at a local  
10 maximum when there is a much larger global maximum still to find.

Once the best possible vector set according to the annealing algorithm has been achieved, then the vectors are analysed, in a vector categorization unit 24 in the light of the results shown and the process. Those vectors having the highest numbers of results tend to represent the  
15 steady state region of the process. Those vectors having the lowest numbers of results tend to represent undesirable states in the process. Vectors in between the two extremes often represent states in which minor changes could usefully be made to the input values in order to better maintain the steady state. The vectors having the lowest numbers of corresponding  
20 results may thus be associated with alarms, demanding immediate action to be taken in the process. The intermediate vectors may be associated with advice given to the process manager or minor tweaks to the process. The



steady state vectors may be associated with very minor tweaks depending on the associated input variables.

An advantage of the use of the vectors is that although a certain overall result may be perfectly acceptable, the vector may easily show that a certain input value is heading out of line and is being masked by other input values compensating for it. This is a situation which is hard for a process engineer to spot but which the vector model will reveal quite easily.

More generally, the vectors represent the inputs that gave rise to any given output produced by the process. In the prior art it was necessary to see a perturbation in the output and from that to deduce that there was something wrong and then use a combination of experience and guesswork to decide which input to change to correct the problem. With embodiments of the present invention however, an automatic association is drawn up between a received output and the inputs that are likely to have given rise to it. Thus the model is able to deduce that a certain input needs correcting even if the overall result looks totally acceptable.

In particular the model preferably provides an analysis of a process involving multiple inputs in terms of all of the inputs in an empirical manner. In prior art systems, only the behavior of one or at most a small number of inputs was effectively accounted for and in general, it was not possible to see when the effect of one input was being masked by another.

Reference is now made to Fig. 2, which is a generalized flow diagram of the system shown in Fig. 1. In Fig. 2, a series of input parameters are identified as being significant in effecting the output of a process. Each input parameter has an expected range. Each expected range is discretized into a series of subranges and a vector is built for each possible combination of subranges. The process is then monitored to obtain a statistically significant set of samples, each sample comprising a process output and the inputs corresponding thereto.

Each output is then attached to the vector that corresponds to its inputs, so that at the end of the sampling period many thousands of samples have preferably been taken and at least most of the vectors have a set of results associated therewith which are statistically analyzable. The vectors are annealed based on the results of a statistical analysis, as described above.

Reference is now made to Fig. 3, which is a generalized block diagram showing how a model derived as described above in respect of Figs. 1 and 2 may be used for monitoring and control of a process.

In Fig. 3, an output measuring device 30 obtains an output measurement  $y$  from the process. A vector identifier 32 relates the measured  $y$  to the total output space and finds the vector  $v_1..v_n$  that best describes that output. The vector is associated with some state of the process, indicated by a label  $s$  attached to the vector. The label is analysed by an instruction finder 36 and if it requires an action to be performed then

an action processor 38 carries out the action. The action may, for example, be to set off an alarm and halt the process immediately, inform the supervisor that a certain input needs correcting, or automatically modifying a process input, or even simply provide a status report.

5           Reference is now made to Fig. 4, which is a generalized flow diagram showing the system of Fig. 3. As shown in Fig. 4, an output of the process is measured. The measured output is associated with a corresponding vector, and any action associated with the corresponding vector is then carried out as necessary.

10           Reference is now made to Fig. 5, which is a flow diagram showing how a model of the type described above may be used as an aid to understanding a process. In the embodiment of Fig. 5 a process is first identified for study. An output of the process is then identified. The process under study may be a part of an overall process, but it should have  
15           an identifiable output.

Once an output has been identified, then all parameters that could possibly affect the output are identified. This could for example be assisted by building a knowledge tree.

20           The process is then empirically monitored and a series of vectors are built up using the procedure of Figs. 1 and 2. The vectors may then be analysed to indicate which are important inputs to the process, whether any inputs are irrelevant, and what actions can be associated with given inputs

to better manage the process. If the model fails to converge, that provides a good indication that a significant parameter has been omitted.

For example a control engineer in charge of a process for manufacturing semiconductors may feel that a significant factor is not being taken into account. A knowledge tree is built up indicating all factors present in the manufacturing environment. Use of the above method allows empirically determined values to be assigned to each node of the tree. Analysis of the tree may then for example indicate that the missing factor is background room temperature. Once the tree has been created using the above method, it is then possible to use intelligent decision-making to decide, based on the tree, what corrective action to take.

As a further example, if the process is being run on several machines in parallel, the machine used may be found to be a variable. Analysis of the tree may be able to identify that certain other parameters behave differently on the different machines. Thus vectors indicating a certain quality of input material may be associated with good output values with one of the machines and worse output values on another machine. Intelligent decision making may indicate that input material can be assigned to the different machines on the basis of its quality.

When programming intelligent decision-making it is preferable to classify variables as easy to change, difficult to change and beyond the possibility of control. Also, it will often be the case that monitoring will be

confined to a part of a process but that the method will indicate that a change is needed to a previous stage in the process.

It will be noted that whereas most of the description above has described continuously variable parameters which are then discretised, a  
5 "machine used" parameter is already a discrete variable and can be incorporated directly into vector formation.

As will be indicated in the examples below, the process need not be restricted to the field of industrial manufacture. The method of Fig. 5 is applicable to any situation in which an output can be analysed in terms of a  
10 plurality of inputs.

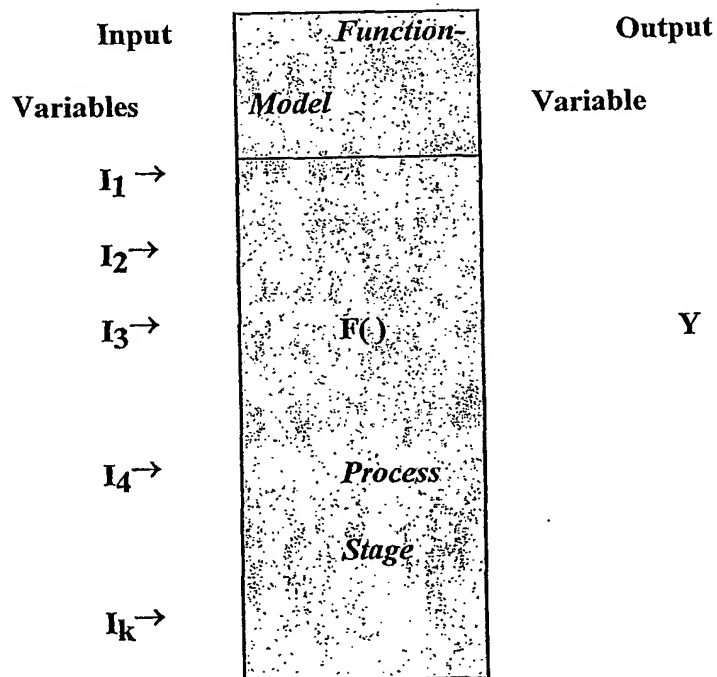
An example of a non-industrial application in which a correct analysis of the data requires careful relating of the outputs to the individual inputs is a program to advise people regarding body weight. The use of a person's body weight as a basis for a medical recommendation is likely to  
15 fail unless the weight is effectively correlated with age, height, sex and other parameters before being associated with medical outcomes.

An embodiment of the present invention may provide a process output empirical modeler (POEM) which can be utilized to define an empirical relationship between measured input value(s) and output value(s)  
20 of a parameter or parameters associated with a single process stage.

Enlarging on what has been described in respect of Fig. 1, by denoting a given set of measured input values of a measured parameter as  $I = \{I_1, I_2, I_3, \dots, I_k\}$ , and a resulting output measurements of the same

parameter as Y it is possible to calculate a functional relationship between an input and output value of such a parameter. This can be achieved by the following:  $Y = F(I_1, I_2, I_3, \dots, I_k)$ , wherein F represents a function which is determined according to the teachings of the present invention as is further exemplified hereinunder and which enables to predict an output value Y, at the end of the stage, from the values of the input variables (see Scheme 1).

**Scheme 1**



10

F is generated by processing actual measured parameter values and employing non-parametric statistics. The resulting functional relationship describes the behavior of a process stage and can be used for

modeling of the process stage, thus allowing simulation, prediction and process control.

#### A Basic Algorithm:

5            Assuming that the measurements of each input parameter,  $I_j$ , in a process varies within a known interval, based on actual data, one can divide this interval into a number of sub intervals. As is further exemplified herein, the measurements of parameter  $I_j$  is classified according to sub intervals and is thus presented and treated as discrete variables. The actual  
10    method of interval division into sub intervals, and the number of sub intervals thereby formed, may be left to the discretion of the skilled person. Therefore, for reasons of clarity and without any intention of loss of generality, assume that the interval of values of parameter  $I_j$  is divided into three sub intervals of equal length, denoted by  $A_j$ ,  $B_j$  and  $C_j$ . Thus, each  
15    individual measurement is classified to either the  $A_j$ ,  $B_j$  or  $C_j$  sub intervals and a measurement array of all  $k$  input values of the process stage is represented by a  $k$ -tuple, in which each entry assumes one of the values  $A_j$ ,  $B_j$  or  $C_j$ .

          For example, assume a function of 4 variables ( $k=4$ ) and for all  $j$   
20    as follows:

$$0 \leq A_j < 10, 10 \leq B_j < 20, 20 \leq C_j \leq 30$$

and an array of the input measurements (or input vector), corresponding to the output measurement 17.40, is equal to (12.00, 5.56, 23.20, 3.00).

Omit the index  $j$ , and denote the first interval by  $A$ , the second by  
5  $B$ , and the third by  $C$ .

In a functional notation:  $17.40 = F(12.00, 5.56, 23.20, 3.00)$ .

In this case, the discrete vector  $[B, A, C, A]$  is associated with the value 17.40.

Construct a discrete function  $FD$ , which accurately represents the  
10 non discrete function  $F$ , provided the number of sub intervals is sufficiently large.

The discrete function  $FD$  assumes in this case exactly  $3^4 = 81$  different discrete vectors. Any measurement input vector (which in this case is a 4-tuple) is classified in this case, to one of a finite number (81) of  
15 possible discrete vectors  $[A, B, C, A]$ ,  $[B, A, C, C]$ , etc.

Now, take a large number of input vectors, each corresponding to a measured output, and translate each vector to the corresponding discrete vector. The different discrete vectors will typically appear many times in the list. For example, one may obtain  $n$  repetitions of the discrete vector  $[B,$   
20  $A, C, A]$ , each corresponding to an output  $Y_i$ . Similarly, for each of the 81 discrete vectors there will be a set of outputs.



Define the value of the discrete function FD at  $[B, A, C, A]$  as the average of  $Y_i$ , provided that certain statistical criteria, which is defined below, are met.

Thus  $Y = FD[B, A, C, A]$ , where  $Y$  is the average of  $Y_i$ . The  
5 standard deviation (SD) of the  $Y$  values is recorded for each discrete vector.

It is useful in some cases to record the whole distribution of the  $Y$  values, corresponding to each of the discrete vector.

Reference is now made to Fig. 6. Fig. 6 represents six input variables, which may each be divided into three discrete regions labeled A,  
10 B, and C. Thus the input space may be defined by a series of resulting input vectors which may be denoted BACCCA, BCCABC, etc. The corresponding output distributions differ in shape, size and location. If an output is to be defined between upper and lower specification limits (USL and LSL respectively) suitable response distributions are selected.

15 Repeating the above mentioned process a finite number of times defines the discrete function FD. Using FD as a discretization stage, a continuous function (model)  $F$  is generated.

As described above, each sample of data can be described as a distribution with a mean, a range and a standard deviation within predefined  
20 upper and lower limits. Each sample may comprise values of two or more input parameters, and in the present illustration six such parameters are shown. Thus, as shown in the right hand side of the diagram an overall

distribution can be defined as the sum of a plurality of separate parameter distributions.

The embodiment examines the input variables which lead up to the overall distribution, and, as a result of such an examination, is in a better position than the prior art to understand the overall process, since the prior art relates only to the overall distribution. Thus a specific one of the various input parameters can be identified as being responsible for variations etc. and this knowledge can be used, for example to decrease variability in the output.

The inputs themselves may be of a continuous nature, and in order to process them they are divided into discrete components, herein labeled A, B, and C. Vectors are formed for each of the possible letter combinations for the six inputs and any input received is assigned to the appropriate input vector.

Reference is now made to Fig. 7, which is a simplified tabular diagram showing a series of inputs and how they may be discretized. A series of inputs 5 to 8 are each related to an output 9. Each input is limited to a certain range but is otherwise continuous within that range. In the table each input range is divided into four sectors A to D.

Reference is now made to Fig. 8 which shows a series of vector values, all related to a single input, wherein each vector has a full set of statistical values associated therewith. The vectors are used to form a lookup table for interpreting measured output values, the vector value.

### Function Determination

In the basic algorithm described above, the relationship  $Y = FD[S1, S2, S3, S4]$ , where  $[S1, S2, S3, S4]$  is any one of the 81 discrete vectors, was defined using  $Y = \text{average } y$ , the average of measurements of output values corresponding to the vector.

In this section, a criterion of reliability of  $Y$  will be defined and sufficiently large number of repetitions ( $n$ ) will be determined, such that the estimation of  $Y$  as an average will be accurate (under some defined criteria).

Taking the empirically measured outputs corresponding to each one of the vector (out of the 81 possible cases) as a random sample of  $Y_j$ , where  $1 \leq j \leq 81$ , it is preferable to test the standard deviation or like statistical parameter among 81 sample means, or equivalently to test the null hypothesis that the sample means are practically equal. A suitable statistical tool for this test is the "Analysis of Variance" (ANOVA).

The first indication of the prediction capability of  $FD[S1, S2, S3, S4]$  is expressed by applying the ANOVA test to the different output means corresponding to the 81 discrete vectors. This will indicate whether a move from one vector to another vector yields a change in the value of the function  $FD$ . This is a necessary condition for a predictor  $FD$ . For each average  $Y_i$  one calculates also the variance  $\sigma_i^2$  and a p-value  $p_i$ .

In some cases, several different discrete vectors will correspond to the same output value. This is because, mathematically,  $FD$  is not

necessarily a one-to-one function. In those cases, the average  $Y_i$ , the variance  $\sigma_i^2$  and p-value  $p_i$ , will be calculated for each cluster. Further, t-tests can be performed for any pair of clusters, to examine the hypothesis that two clusters means are equal. FD will have statistical significance if  
5 this hypothesis is rejected for any cluster pair.

### Algorithm evolution

#### Function fine-tuning:

Given a 4-structure, the specific division of each of the intervals  
10 (above, each interval was divided into equal sub intervals), as well as the number of sub intervals (above, the number was three), has in general an impact on the variance  $\sigma_i^2$  and the p-value  $p_i$ . Thus, cluster grouping and the related pairwise t-tests should also yield different results. Hence, the function's predictive quality may be improved. Using iterations of the  
15 algorithm, one applies the ANOVA test for different divisions of the intervals, in order to get lower values of the variance -  $\sigma_i^2$  and the p-value  $p_i$ .

#### Elimination of Redundant Input Variables:

20 Although a comprehensive set of variables (I) may affect Y, some variables in the set I which have no effect on Y are redundant and thus could be eliminated; in other words, Y may be a function of a subset of I.

Thus, those variables in the set I that are redundant, are preferably eliminated such that the algorithm described above is applied to the most concisely effective set of input variables. This stage could be carried out by a number of different well-known algorithms, such as, but not limited to  
5 Factor Analysis and Principal Component Analysis, both widely used in conventional statistics.

#### If-Then Rule Learning:

The function F described hereinabove has a continuous range,  
10 meaning that Y can assume any value in a given interval. If the range of Y's is divided into sub-intervals, an "If-Then" rule can be applied to the data as is further detailed hereinbelow.

#### Applications

15 The functional relationship described above can be applied to any process which includes one or more stages and which, for each of the stages, receives an input and produces an output. The use of the logical-mathematical model(s) generated according to the teachings of the above-described embodiments enables variability detection at a sensitivity level  
20 which far supersedes that achieved by prior art statistical models and as such greatly contributes to performance improvement and optimization of any process.

The following example describes the method of the present invention as applied to the semiconductor industry, which at the present uses very advanced and sophisticated process in terms of data availability and accessibility.

5 Semiconductor manufacturing:

Reference is now made to Fig. 8, which shows a stage in a wafer production process, which can usefully be monitored and controlled using a model according to the present invention.

10 In wafer production, a chemical mechanical polishing (CMP) process is used for polishing and removing an oxide layer from a wafer surface. In such a process, which is shown schematically in Fig. 8, it is essential to maintain a planarized wafer surface for processes which follow CMP. It will now be demonstrated how the algorithm of the present invention is used to optimize this process and reduce the final thickness  
15 variability of the wafers produced.

Initially, the CMP process is analyzed taking into account the various variables and the interactions therebetween.

Following analysis, a valid model which represents and qualifies the process interactions is constructed by running a set of experiments in which  
20 raw data is analyzed and thereafter utilized to generate a model.

An embodiment of the invention during a CMP process is realized in Figs. 8-13. In the CMP process 800, which is shown in Fig. 8, a wafer is

sequentially polished using two rotating platens; platen 1, 801 and platen 2, 810. In Fig. 8 arrows designate all the inputs and outputs of the process.

The measurable but uncontrolled inputs for the polishing process using platen 1 801 include the incoming wafer thickness 803 and pad life 802, i.e. the amount of time which the polishing pad is already in use at that platen.

The controllable inputs of the process are the retaining ring pressure 805, i.e. the pressure in which the wafer is pressed towards the polishing pad, the platen rotating speed 806 and the polishing time 804 of the wafer.

The thickness of the out coming wafer is an output parameter of the first polish stage performed at platen 1, 801; However it also serves as a known (either measured or calculated) input 807 for the second stage of the polish process which is performed at platen 2, 810.

The second stage has different values for its polish pad life 812, wafer's polish time 814, retaining ring pressure 815 and platen speed 816.

The wafer after this stage is characterized by its final thickness 820 and uniformity 818. These outputs are measured for each of the out coming wafer together with the corresponding input parameters at the second stage, that caused these outputs. All these values are tabulated in a raw data table 900, which is shown in Fig. 9.

In Fig. 9, it was assumed for the sake of simplicity of explanation, that all the incoming wafers have a nominal constant thickness, thus the

effect of the incoming wafer's thickness was ignored and its respective input did not appear.

The entries in each of column (field) of raw data in table 900 are the values of the input which that field consists of. As for inputs, the fields (columns) include values for the retaining ring pressure 905, the platen rotating speed 906, the pad life 902 and the polish time 904. As for the output parameters the fields include the uniformity score 918 and the removed thickness 919.

Thus, for an actual experiment (polishing a single wafer), the raw data is represented in a record (a certain row), e.g. row 920 of table 900.

Each input's range, i.e. the interval between the maximum and the minimum values of an input in a column, is then divided into subranges according to the teachings of the invention which was taught in connection to the formation of the discretization table that is shown in Fig. 7, so that each input can be represented by one of N levels of discrete parts.

This is shown in Fig. 10 for both stages of the CMP, where the various process inputs have the same notation as in Fig. 8.

In Fig. 10, N=5 (A to E) and a combination of the respective levels, a level per input, generates a vector e.g. vector DBBCD 1020 or vector ADEDB 1030.

Suppose that fine tuning was accomplished; e.g. that the boundaries of the subranges of each input parameter are such as to produce the most distinctive outputs. The next stage in this embodiment of the invention is



to further provide raw data collected during a CMP actual job, to assign an input vector and an output result to each of the polished wafers, and to establish a lookup table of all the vectors and their respective output values; an example for such a lookup table is shown in Fig. 11.

5           Fig. 11 shows a look up table 1100 for the first stage which is performed at platen 1, 801 of the CMP process.

In Fig. 11, each of the vectors appearing in column 1101 of lookup table 1100 represents a CMP process setup in which each of its inputs is confined within its respective subrange. The resulting average wafer's thickness (vector value) and the standard deviation (sigma) for a sample  
10           population of n wafers which are included in a certain vector e.g. vector FFFFFF 1106 of table 1100, are shown in column 1103 and 1104 of table 1100, the number of n is given in column 1102 of table 1100.

One would like to relate the values in columns 1103 and 1104 for  
15           each vector-j (a sample which include  $n_j$  members) to the respective thickness mean and variance of the distribution of the real (entire) population of wafers which are the yield of a process having the input's values of that vector.

These statistical relations are represented in Columns 1105 and 1106 of  
20           table 1100, which show the mean and the sigma standard error respectively.

Feed forward, 821 in Fig. 8 is accomplished by a continuous on line monitoring of the values of the measurable inputs of the process for each wafer and dynamically assigning the controllable input accordingly to form

the preferred vector having the desired output. As a result the setup of a polishing machine e.g. its platen rotation speed or its retaining ring pressure can be subjected to automatically changes each time a new wafer arrives.

This is an innovative routine in particular with regard to the first  
5 stage thickness output 807, which is fed forward as an input to the second stage.

Feedback, 822 in Fig. 8 is performed whenever one gets results which are off of target values, thus he has to shift toward process setup which is included in a different vector that according the look up table will  
10 divert the results into the target region.

For example, suppose a range of  $4000 \pm 200$  in wafer's final thickness is desired and one gets an unallowable thickness spread while working at an input setup which corresponds to vector BBBDD in Fig. 8, using the discretisation symbols of Fig. 7.

15 According to lookup table 1100, spread can be improved by changing the input setup into one of other three possible setups according to input vectors which are: BBBBC 1007, BBBC 1008 and BBBCD 1009.

In this case the vector BBBCD 1009 is clearly preferable because among the three candidates, it has the lowest combination of sigma standard  
20 error with the mean standard error, thus it represents the most stable working envelope.

Fig. 12 momentarily shows a window of a window-based graphical interface during the on line computerized monitoring and control of the CMP process according to the present invention.

Reference is now made to Fig. 13 referenced as 1200, which shows  
5 two histogram charts 1210 and 1220 of a CMP processes according to the present invention. The upper process 1220, being the one in which the teaching of the present invention as described in this example was used to perform a closed loop process, while the lower chart 1210 represents an open loop process. The moments and capability analysis associated with  
10 each chart have their usual statistical meaning as accepted in the art of process control.

As it is evident from histogram charts 1220 and 1210 and their analysis, the CMP process in the closed loop process was improved by 50 % in sigma terms compared to the open loop one, using the system and the  
15 algorithm of the present invention.

#### Dynamic order routing-financial applications:

Moving away from manufacturing processes, an example follows of how embodiments of the present invention may be applied to a transaction system in order to assist brokers to fulfill their various legal  
20 requirements to obtain the best possible deal for their client, from any one of a number of sources offering the required security.

There is a considerable variety of trade execution points, including exchanges (e.g. NYSE, AMEX), regional exchanges (e.g. BSE, PHLX) ECN's (Electronical Communications Networks e.g., Redibook, Instinet, SelectNet) and over the counter (OTC) market makers (Fleet Trading, Knight). Once a trade order has been accepted there is a need to  
5 determine the execution destination for that trade order. This process is known in the art as "order routing".

The traditional approach for order routing is to use a pre-defined rule-based system. This approach utilizes attributes from the order (e.g.,  
10 order size) and the security being traded (e.g., non-listed, listed etc.) to determine a routing destination. The main disadvantage with this approach is that it does not take into consideration many dynamic factors such as volatility and liquidity which change with time and market.

Another approach for order routing is known as "dynamic  
15 routing". This approach uses real time data from the possible execution points in order to find the best route of execution for a certain order. The use of dynamic routing can yield significant benefits to a client placing an order.

The term "best execution" is determined largely by the price of  
20 the execution and the opportunity for price improvement. However, there are other factors such as speed of execution and likelihood of execution that may be equally important.

It should be emphasized that the term "best execution" is not only an economic goal but also a legal obligation of the brokerage firms. According to the SEC and NASD rules, a member must use reasonable diligence to ascertain the best inter-dealer market for the subject security and buy or sell in such market so that the resultant price to the customer is as favorable as possible under prevailing market conditions. Thus, the quality of execution must always be viewed from the customer's perspective and not that of the firm.

The algorithm of the above-described embodiments can be utilized for solving problems associated with order routing. First, each execution destination is identified and analyzed for measurable inputs that affect parameters that determine the quality of an execution (e.g., price, speed, likelihood, etc.) These measurable inputs can be, for example, liquidity (measured, for example, by the bid/ask imbalance), volatility (measured, for example, by the spread size), current price (relative to the lowest price), order size and time of the day. Once this information is acquired, past data can be utilized to build an execution destination for each stock and to build a predicted distribution of each of the outputs for every set of measurable inputs. In other words, one can create a lookup table for predicting the results of sending an order to a certain destination. This table can be utilized for optimizing a destination for an order. This optimization can be done in two ways as follows:

**Feed-Forward** - collecting relevant real time data from all possible destinations of a new order, using the lookup table to compare the predicted output and sending the order to the optimal destination according to the results obtained from the lookup table.

5           **Feedback** - when an order which is sent to a certain destination produces a result which is significantly worse than expected, the information is assimilated in order to correct 'behavior' for future orders.

10

#### Weight Monitoring:

Assume a system whose task is to monitor weights and to detect overweight or underweight trends in the population. Our system accepts weight measurements and will identify outstanding items. For simplicity  
15   assume that weights higher than 95 kg. And lower than 55 kg. Are considered outstanding.

Our system will react to a 100 kg. measurement. However, if the relevant person's height is 205 cm. then the alarm will be a false one, since for a very tall person a weight of 100 kg. is normal and healthy. Thus a  
20   better monitoring should relate to the weight as function of height and look at weight distributions per height, or height subinterval. A weight of 70 kg. will pass the system unnoticed, but if the relevant person's age is 5, the the person is definitely overweight. Therefore weights should be treated as

functions of two variables; height and age. Similarly we should consider parameters as sex, ethnic origin etc. as effecting weight.

Monitoring the weight as a multivariate function will yield a more sensitive monitoring while reducing false alarms.

5        We may use the present invention to create the relevant weight distributions from data, and rather than monitoring the population by the entire population distribution use the relevant specific vector distributions.

#### Health Care Applications:

10        Medical databases contain information which is reflective of empirical medical results and as such probably contain information and relationships not known at the present to medical science because to date there has not been the tool to effectively take into account the effects of multiple inputs in a comprehensive and systematic manner.

15        Artificial Intelligence (AI) can be used to extract knowledge of medical significance from such databases. The algorithm of the present invention can be utilized in data mining techniques to determine a relationship between causes (Input values) and effects (Output value) and to functionally model such relationships. Models generated can be applied to  
20        improve medical decision making capabilities.

### 1) Treatment of simultaneous multiple pathologies

Treatment of a specific pathological disorder in an individual may effect other disorders. The optimal selection of treatment in the case of multiple disorders may be complex, since it depends on many parameters and interrelationships. The algorithm of the present invention can be used to model multiple disorder situations and as such to improve decision-making capabilities. In this connection reference is now made to Fig. 14 which shows how a variety of tests combined with a patient history can be measured and compared with an output in terms of the likelihood of a given pathology. Provided the input sample is sufficiently large, useful information may be obtained concerning predictions of the likelihood of the given pathology, in the same way as useful indications were gathered above in relation to process input-output relationships in silicon wafer manufacture.

Reference is now made to Fig. 15, which is a simplified diagram of a model showing various inputs including diagnosed conditions and applied treatments, being related to a series of outputs. It will be appreciated that in some cases, selection of optimal treatment may be beyond the capabilities of a physician due to the large number of factors, their complexity, the interrelationships therebetween and the minimal time available for decision making. In cases of three or more simultaneous illnesses, decisions will rarely be optimal, resulting in suboptimal patient care and undue expenses resulting from unneeded treatments.



For example, HRT can be utilized to lower the incidence of heart dysfunction, but tends to raise blood sugar and triglyceride levels. Beta-blockers can alleviate hypertension but have deleterious effects on coughing and asthmatic illnesses. In such cases, it is oftentimes difficult for a physician to decide what course of treatment to apply which would result in lowest hospitalization rate, doctors' visits and lowest treatment cost. As shown in Figure 6, the algorithm of the present invention can successfully map these complex relationships and indicate for a given combination of disorders, the best possible treatment regimen.

## 2) Analyzing lab tests - "If-Then" rule learning

Experienced physicians can qualitatively relate selected laboratory tests with a pathological condition and thus indicate the presence or absence of such a condition. The present invention enables to quantitatively relate laboratory tests to pathological conditions by generating a quantitative table (function) from an extensive database containing the lab results and the respective pathology occurrence.

As shown in Figure 14, for example, the quantitative table can be used along with a logical set of rules in the following manner: if the result of Blood Test 1 is high and the result of Blood Test 2 is low and the result of Blood Test 3 is medium and the result of Blood Test 4 is high, then, considering the patient records, the likelihood of the pathology is high.

The high/medium/low levels are only examples; one may define additional grades, such as, for example, "very high", "high-medium" etc.

The algorithm of the present invention generates the functional relationship between blood tests (input) and conditions (output) by utilizing actual data and non-parametric statistics during a "learning period" in which collected data or stored data is used to generate and calibrate the function.

### 3) Individualization of treatment

Yet another health related application of the present invention relates to individual customization of treatment, drugs, drug doses etc. By accessing patient records and utilizing patient characteristics as the input values and the recorded success of the treatment as the output, the algorithm of the present invention can optimize treatment according to patient parameters.

An embodiment of the present invention uses the boundary value ranges assigned to input constants and variables to form data vectors for a given stage in a process. Each input constant or variable is a component entry of the vector. Given the inputs depicted in the three diagrams in Figure 16a and their respective boundary values, it is seen that the following 24 data vectors exist for the output produced by the inputs depicted in Figure 10a: { (A1, A2, A3), (A1, A2, B3), (A1, A2, C3), (A1, A2, D3), (A1, B2, A3), (A1, B2, B3), (A1, B2, C3), (A1, B2, D3), (B1, A2, A3), (B1, A2, B3), (B1, A2, C3), (B1, A2, D3), (B1, B2, A3), (B1, B2, B3), (B1, B2, C3),

(B1, B2, D3), (C1, A2, A3), (C1, A2, B3), (C1, A2, C3), (C1, A2, D3), (C1, B2, A3), (C1, B2, B3), (C1, B2, C3), (C1, B2, D3) }.

Referring again to Figure 16a, for the sake of example, assume that 10a.1 represents an input constant, and that 10a.2 and 10a.3 represent input variables at a given stage in a process. The boundary values for 10a.1 are  $x_1 = 24.98$  mm and  $y_1 = 25.02$  mm, where B1 is the preferred boundary value range for values between 24.98-25.02 mm inclusively, A1 is the boundary value range for values less than 24.98 mm, and C1 is the boundary value range for values greater than 25.02 mm. For 10a.2, there is one boundary value  $x_2 = 10.00$  mm. A2 is the boundary value range for values less than or equal to 10.00 mm, and B2 is the boundary value range for values greater than 10.00 mm. For 10a.3, there are three possible boundary values, denoted  $x_3$ ,  $y_3$ , and  $z_3$ . A3, B3, C3, and D3 represent four possible boundary value ranges for the height of an item. The possible range of the height of the item varies from 0.00 mm to 10.00 mm. A3 is the boundary value range for values greater than 0.00 mm and up to and including 2.50 mm, B3 is the boundary value range for values greater than 2.50 mm and up to and including 5.00 mm, C3 is the boundary value range for values greater than 5.00 mm and up to and including 7.50 mm, and D3 is the boundary value range for values greater than 7.50 mm and up to and including 10.00 mm.

Figure 16b illustrates a Data Arrays table of data arrays 5000 for a given stage in a process. The table is composed of a column for the

number of the process run 5001, columns for process input 5002, and a column for a given process output constant 5006. The inputs at this process stage are input constant 10a.1 5003, input variable 10a.2 5004, and input variable 5005. Values for these inputs corresponding to the data vector (B1, A2, D3) are received at the second process run 5007, the  $e$ th process run 5008, the  $e+1$ th process run, and the  $f$ th process run. The value for the given output constant for the second process run is O2 5011, the value for the given output constant for the  $e$ th process run is Oe 5012, the value for the given output constant for the  $e+1$ th process run is Oe+1 5013, and the value for the given output constant for the  $f$ th process run is Of 5014; also  $i$ th Inputs 5002: input constant "10a.1" 5003, input variables "10a.2" 5004 and "10a.3" 5005 and their respective Output values 5006 make up a data array for any given Process Run(s) 5001.

If we received values of 25.01 mm for 10a.1, 9.98 mm for 10a.2, and 8.00 mm for 10a.3, this data corresponds to the vector (B1, A2, D3), according to the assigned boundary values. Referring to Figure 10b, assume that the process is executed  $n$  times, and that after assigning boundary values to the data received for 10a.1, 10a.2, and 10a.3, values corresponding to the vector (B1, A2, D3) are received for process runs 2,  $e$ ,  $e+1$ , and  $f$ , where  $e$  is an integer whole number greater than 3 and  $f$  is an whole number integer greater than  $e+1$  and less than or equal to  $n$ . The values O2, Oe, Oe+1, and Of represent the output values received for a given output constant for process runs 2,  $e$ ,  $e+1$ , and  $f$  respectively at the

given stage in the process. The data received for any given process run, such as the value for input constant 10a.1 at run 2, the values for input variables 10a.2 and 10a.3 at run 2, and the value O2 for the output at run 2, are referred to as a data array.

5                   Figure 16c (generally referenced as) 6000 illustrates a sample vector in a vector look-up table. The table is composed of columns for the data vector 6001 and columns for the given output constant data 6002. The entries of the vector for the sample vector depicted in this table are input constant 10a.1 6003, input variable 10a.2 6004, and input variable 10a.3 6005. The types of output constant data recorded in this vector look-up table  
10                   are Average 6006, Standard Deviation 6007, and Population 6008. The sample vector is vector (B1, A2, D3) 6009. The average value for this vector for the given output constant is 6010, the standard deviation is  $\sigma(O)$  6011, and the population number is 4 6012.

15                   After a number of runs deemed sufficient by statistical criteria have been executed, the data arrays are sorted according to the data vectors they correspond to, and various meaningful statistical calculations are performed on the output data. For example, in Figure 16b, data arrays corresponding to the vector (B1, A2, D3) were received for process runs 2, e, e+1, and f. In  
20                   Figure 16c the output data for these four process runs is taken and the average and standard deviation of these four output values is calculated. The average value , the standard deviation  $\sigma(O)$ , and the population number

4 are then entered in the vector look-up table in Figure 16c by vector (B1, A2, D3). This output data is used by embodiments of the present invention for optimization of the given output constant. In addition to average, standard deviation, and population number; other types of meaningful statistical calculations are performed on output data by embodiments of the present invention, such as determining the output constant's Process Capability Ratio (Cpk), and the results of these calculations are used for process control optimization of that output. However, for the purposes of illustration, the examples that follow here refer to calculation and use of only standard deviation, average, and population number of output constant data.

In many process control situations, it is understood that not all possible combinations of boundary value ranges for input constants and variables represent actual valid process input. Therefore, for those vector input combinations that represent invalid input combinations for which the given process cannot be carried out, there will be no corresponding output data in the vector lookup table.

The conventions of assigning boundary values to input data and sorting input data into data vectors enable detection of problematic input combinations and detection of input combinations that were assumed to yield output that is out of process specification standards and actually yield output that is within process specification standards. When problematic or unusual input combinations are detected, embodiments of the present

invention provide appropriate system responses. One of these responses is a self-adjusting feature, which automatically adjusts process input that is out of process specification standards to within specification standards. Other system responses include the sending of automated reports to the process engineer, or in more serious cases the sounding of an alarm or even  
5 cessation of process execution altogether.

For example, it is understood that in certain process control situations, certain vector input combinations will represent input combinations for which the current process can be carried out, however it  
10 has been determined from previous history of the given process that the a given input combination is known to yield output which is out of process specification standards, or that the a given input combination contains one or more inputs outside of process specification standards, or that this specific combination of inputs is unacceptable for reasons related to the  
15 given process. An embodiment of the present invention allows the process engineer to program the system carrying out the given process so that if input combinations or output that are considered unacceptable for either of these reasons are received during process execution, the machinery and/or mechanisms carrying out the process automatically correct the input to  
20 within process specification standards. For more serious cases of this nature, this embodiment of the present invention allows the process engineer to program the system carrying out the given process to automatically sound an alarm instead of or in addition to automatic correction, or to even,

automatically halt process execution altogether; or to report an unacceptable input combination or output to the process engineer, or in more serious cases an alarm is sounded or process execution is halted altogether.

Likewise, in certain process control situations, examination of data in  
5 the vector look-up table shows that certain combinations of boundary value ranges for input constants and/or variables which were assumed to yield output that is out of process specification standards do in fact yield output that is within process specification standards. Or, certain combinations of boundary value ranges for input constants and/or variables where one or  
10 more of the boundary value ranges are considered out of the specification standard for that input do in fact yield output that is within process specification standards.

For example, referring again to input constant 10a.1 and input variables 10a.2 and 10a.3, assume that for 10a.1 the boundary value range  
15 A1 is considered out of process specification standards, that for 10a.2 the boundary value range B2 is considered out of process specification standards, and that for 10a.3 the boundary value range A3 is considered out of process specification standards. However, after applying boundary values to the input data and sorting the input data into data vectors according to the  
20 embodiments of the present invention, the resulting output is found to be within process specification standards. Despite this output, which is within specification standards, such a situation still warrants attention, as the given input combination is still considered to be out of process specification



standards. In such a case, an embodiment of the present invention allows the process engineer to program the system carrying out the given process to report input combinations that are out of process specification standards and yet yield output within process specification standards. The input combination can then be analyzed to determine whether the combination constitutes a new and valid set of input or whether the combination constitutes an invalid set of input despite its output yield within process specification standards. The ability of embodiments of the present invention to determine input combinations of this nature with resulting output within process specification standards is a unique feature of the present invention that is unknown in standard methods of process control.

Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims. All publications cited herein are incorporated by reference in their entirety. Citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the present invention.

It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various

features of the invention which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable subcombination.

It will be appreciated by persons skilled in the art that the present invention is not limited to what has been particularly shown and described hereinabove. Rather the scope of the present invention is defined by the appended claims and includes both combinations and subcombinations of the various features described hereinabove as well as variations and modifications thereof which would occur to persons skilled in the art upon reading the foregoing description.

## WHAT IS CLAIMED IS:

1. A method of monitoring a process having at least one input parameter having an expected range and taking a value within said expected range, and at least one output parameter, said output parameter taking a value which is related to at least one of said values taken by said input parameters, the method comprising:

dividing said expected ranges of said input parameters into sub-ranges,

obtaining a series of values for each of said input parameters,

obtaining a corresponding series of values for said at least one output parameter,

associating each sub-range with values of said at least one output parameter corresponding thereto, and

associating each sub-range with one of a plurality of possible states of said process, thereby to monitor said process in terms of said states.

2. A method according to claim 1, wherein there are a plurality of input parameters defining an input space, and comprising the steps of

dividing each parameter into sub-ranges,

building vectors of combinations of subranges, thereby defining the input range by said vectors,

associating measured outputs with a vector describing corresponding inputs, and

associating said vectors with states of said process.

5           3.       A method according to claim 2, wherein said step of associating each vector with corresponding values of said at least one output parameter comprises associating said vector with a single value being a statistically processed result of said corresponding values of said output parameter.

10

4.       A method according to claim 2, comprising the further steps of

identifying the statistical distribution of output values corresponding to at least some of said vectors,

15       modifying at least one of the boundaries of a subrange used in said vectors,

reassigning said output values to said vectors in accordance with said modified sub-range boundaries,

20       re-identifying the statistical distribution of output values corresponding to at least some of said vectors,

and selecting the subranges giving a better statistical distribution according to a predefined distribution criterion.

5. A method according to claim 4, wherein said steps of modifying, re-identifying and selecting are repeated until a predefined finishing criterion is met.

5 6. A method according to claim 4, wherein said predefined distribution criterion is a low mean square distribution.

7. A method according to claim 2, wherein at least one vector is associated with a probability of occurrence.

10

8. A method according to claim 7, wherein any vector associated with a low probability of occurrence is further associated with an alarm.

15 9. A method according to claim 2, wherein said states are grouped into steady states of said process and states requiring corrective action to said process.

10 20 10. A method according to claim 2, wherein said process is a part of a larger process.

11. A method according to claim 2, wherein said process is at least part of a semiconductor wafer manufacture process.

12. A method of modeling a relationship between a plurality of input parameters each having an expected range and an output parameter, said relationship having a plurality of possible states, the method  
5 comprising:

discretizing said expected ranges into a plurality of sub-ranges,

building vectors of combinations of said sub-ranges of said input parameters,

associating each vector with a corresponding value of said output  
10 parameter, and

associating each vector with one of said possible states, thereby modeling said relationship.

13. A method according to claim 12, wherein said step of  
15 associating each vector with corresponding values of said at least one output parameter comprises associating said vector with a single value being a statistically processed result of said corresponding values of said output parameter.

20 14. A method according to claim 12, comprising the further steps of

identifying the statistical distribution of output values corresponding to at least some of said vectors,

modifying at least one of the boundaries of a subrange used in said vectors,

reassigning said output values to said vectors in accordance with said modified sub-range boundaries,

5 re-identifying the statistical distribution of output values corresponding to at least some of said vectors,

and selecting the subranges giving a better statistical distribution according to a predefined distribution criterion.

10

15. A method according to claim 14, wherein said steps of modifying, re-identifying and selecting are repeated until a predefined finishing criterion is met.

15

16. A method according to claim 14, wherein said predefined distribution criterion is a low mean square distribution.

17. A method according to claim 12, wherein at least some of said states indicate actions to be taken.

20

18. A system for monitoring a process having a plurality of input parameters, each taking values within expected input ranges, at least

one output value taking values within an expected output range, and a plurality of possible operational states, the system comprising:

an input value recorder for recording a series of values of said input parameters,

5 an output value recorder for recording a corresponding series of values of said at least one output parameter,

a range divider for dividing said expected ranges of said input parameters into sub-ranges,

10 a vector builder for building vectors of sub-ranges of each of said input parameters,

a first associator for associating each vector with corresponding values of said at least one output parameter, and

a second associator for associating each vector with one of said possible operational states of said process, thereby to monitor said process  
15 in terms of said states.

19. A system according to claim 18, further comprising a statistical analyzer, associated with said first associator, for producing a single value being a statistically processed result of said corresponding  
20 values of said at least one output parameter.

20. A system according to claim 18, further comprising an intelligent decision maker operable to use said vectors to provide numerical



values for nodes in a decision tree of said process, and to make decisions based on desired outputs and on said values.

5           21.       A system according to claim 18, further comprising a vector annealer for annealing vectors based on the statistical range of said associated corresponding values of said at least one output parameter.

10           22. A system according to claim 21, operable to distinguish between inputs which are effective in governing said process and inputs which are ineffective.

          23.       A system according to claim 18, wherein said states are indications of a quality level of said process.

15           24.       A system according to claim 18, wherein at least one of said states is indicative of corrective action to be taken in said process.

          25.       A system according to claim 18, wherein said process is a part of a larger process.

20           26.       A system according to claim 18, wherein said process is at least a part of a semiconductor wafer manufacturing process.

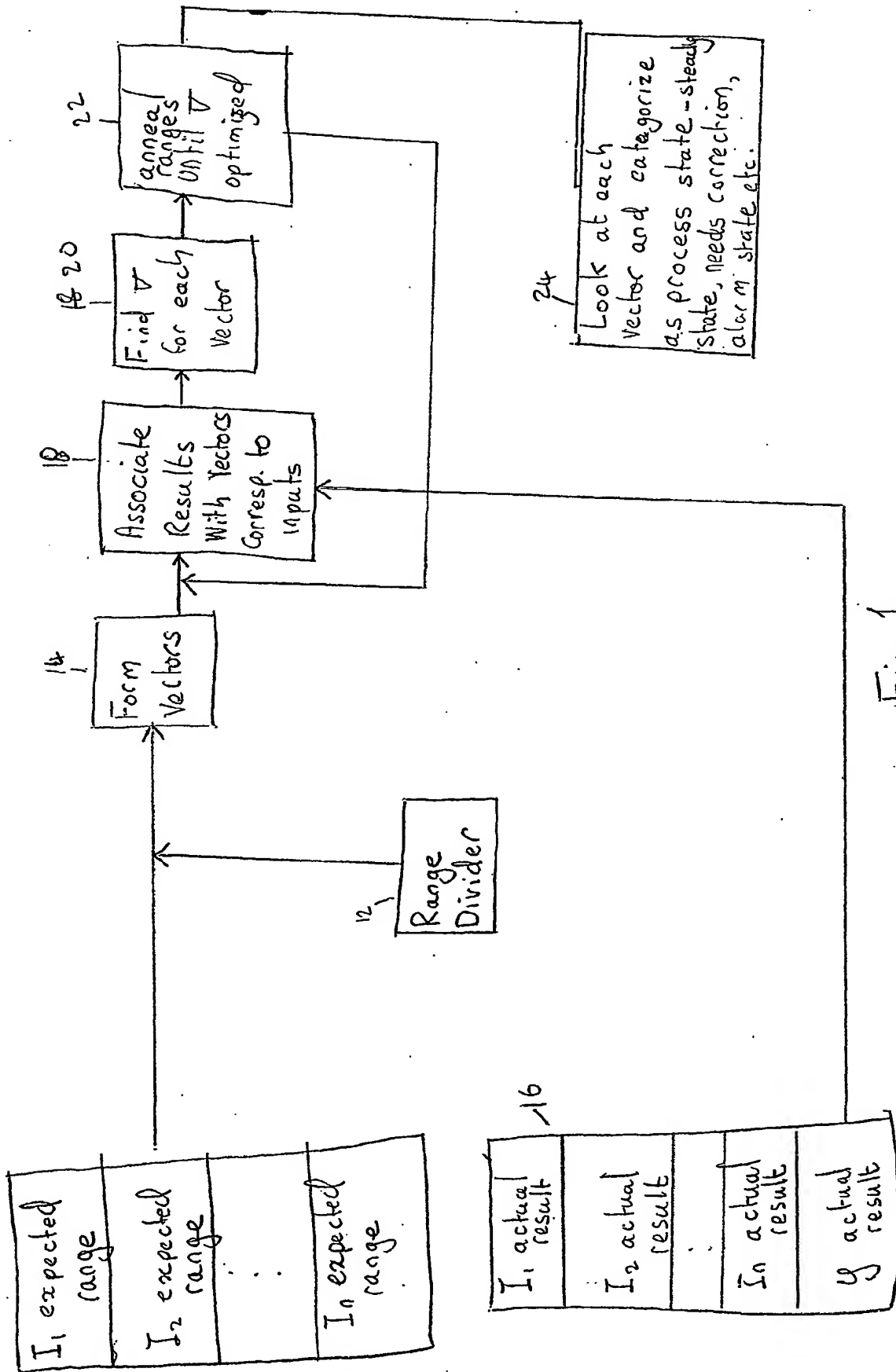


Fig. 1

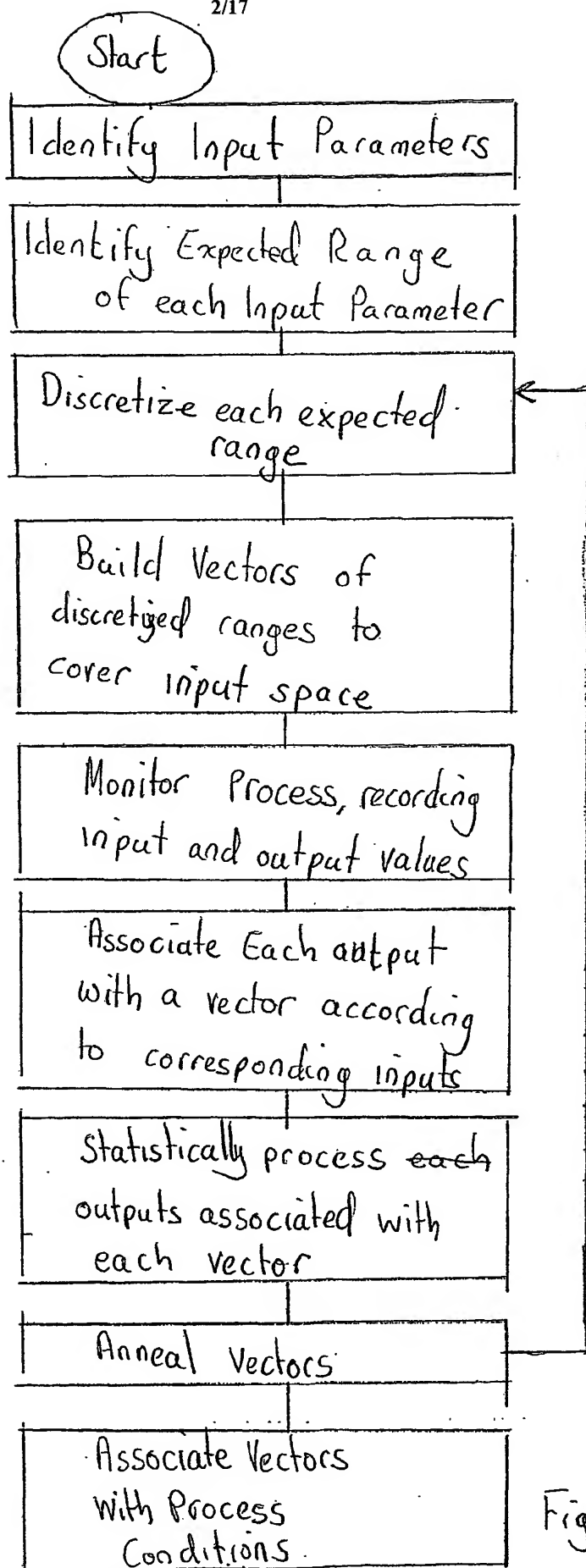


Fig. 2

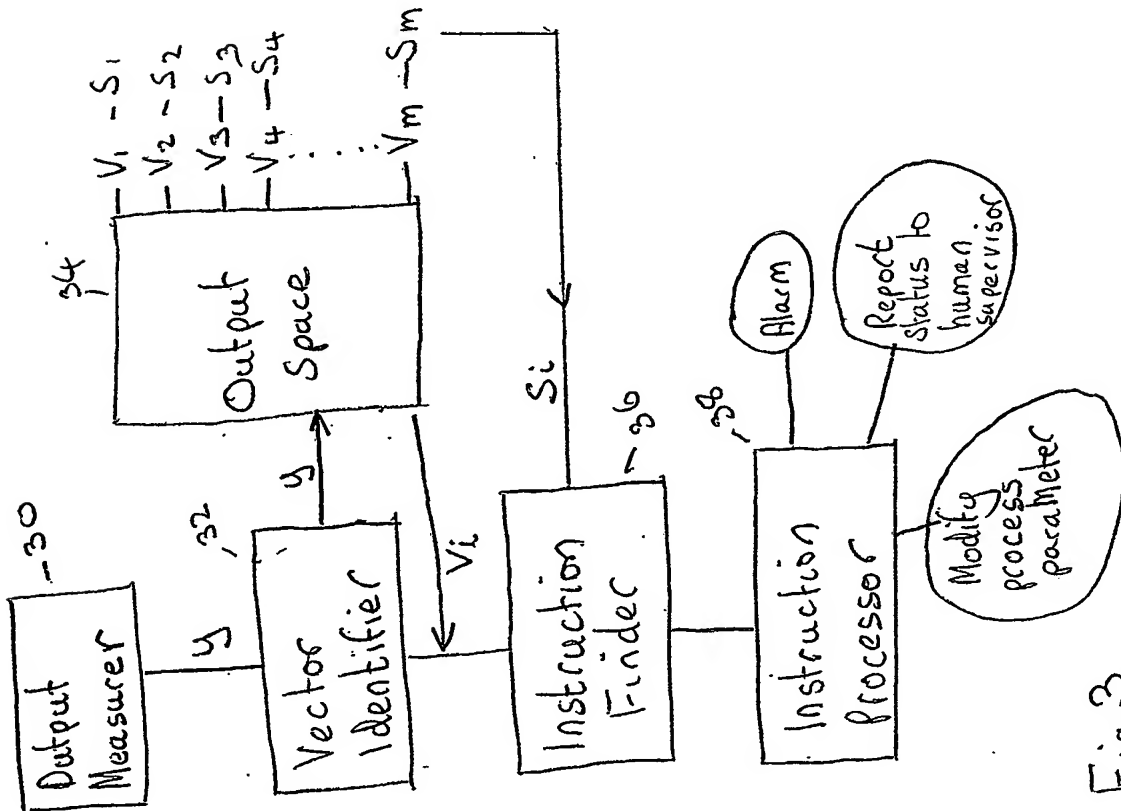


Fig 3

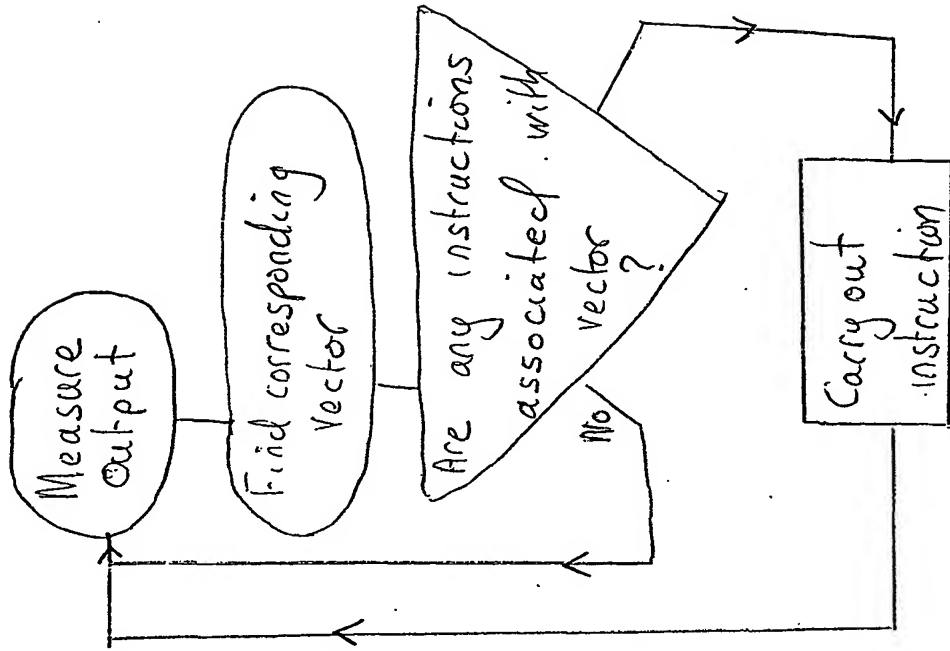
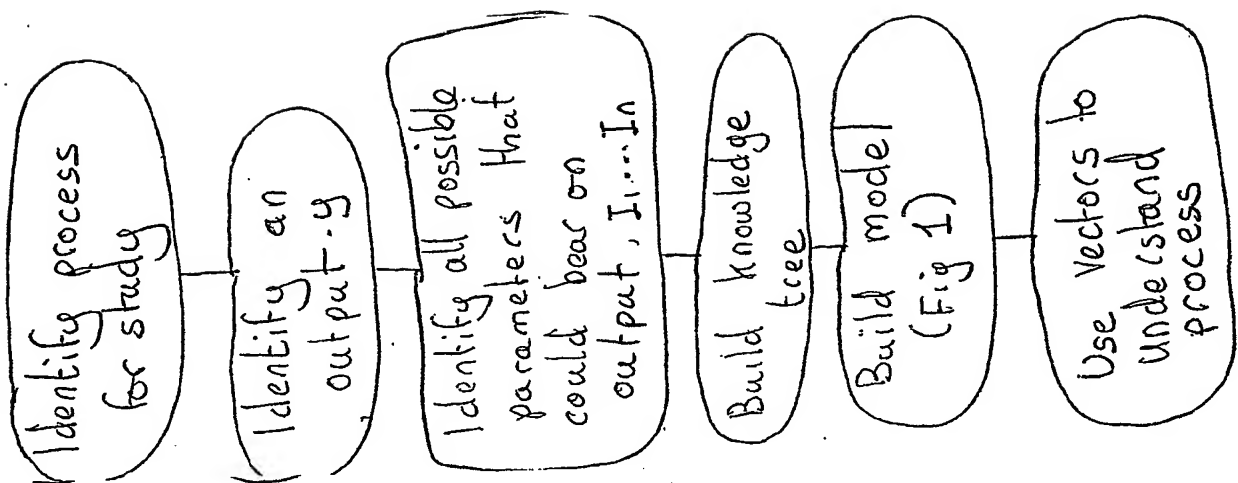


Fig 4

Fig. 5

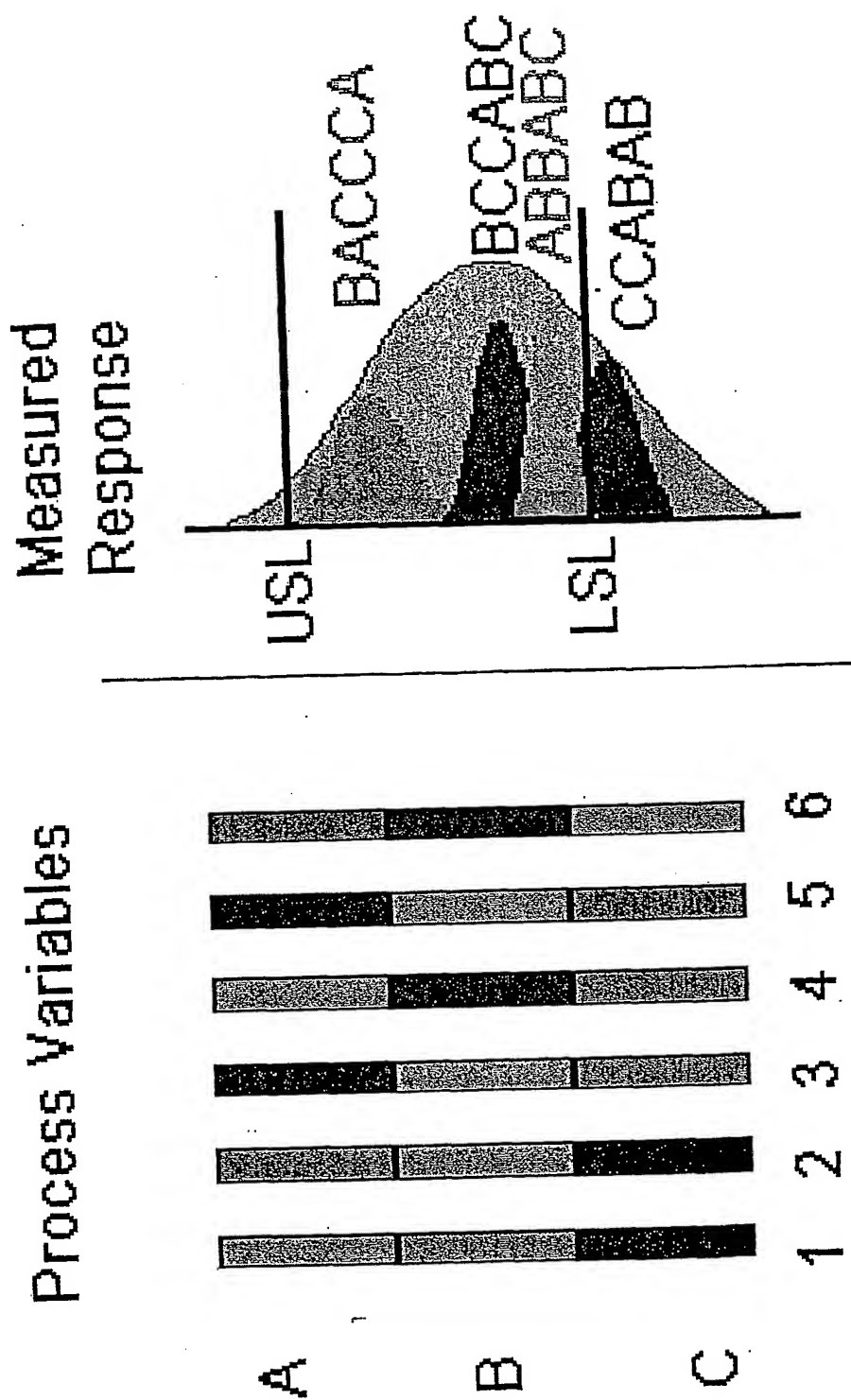


Figure 6

# POEM- Vectors Discretization Table

input_id	output_id	discretization_symbol	discretization_value
5	9	A	2.4
5	9	B	2.5
5	9	C	2.8
5	9	D	3.0
6	9	A	450.0
6	9	B	475.0
6	9	C	525.0
6	9	D	550.0
7	9	A	326.5
7	9	B	335.1
7	9	C	352.4
7	9	D	361.0
8	9	A	230.0
8	9	B	245.0
8	9	C	275.0
8	9	D	290.0

Figure 7

# Control Loop input/output variables

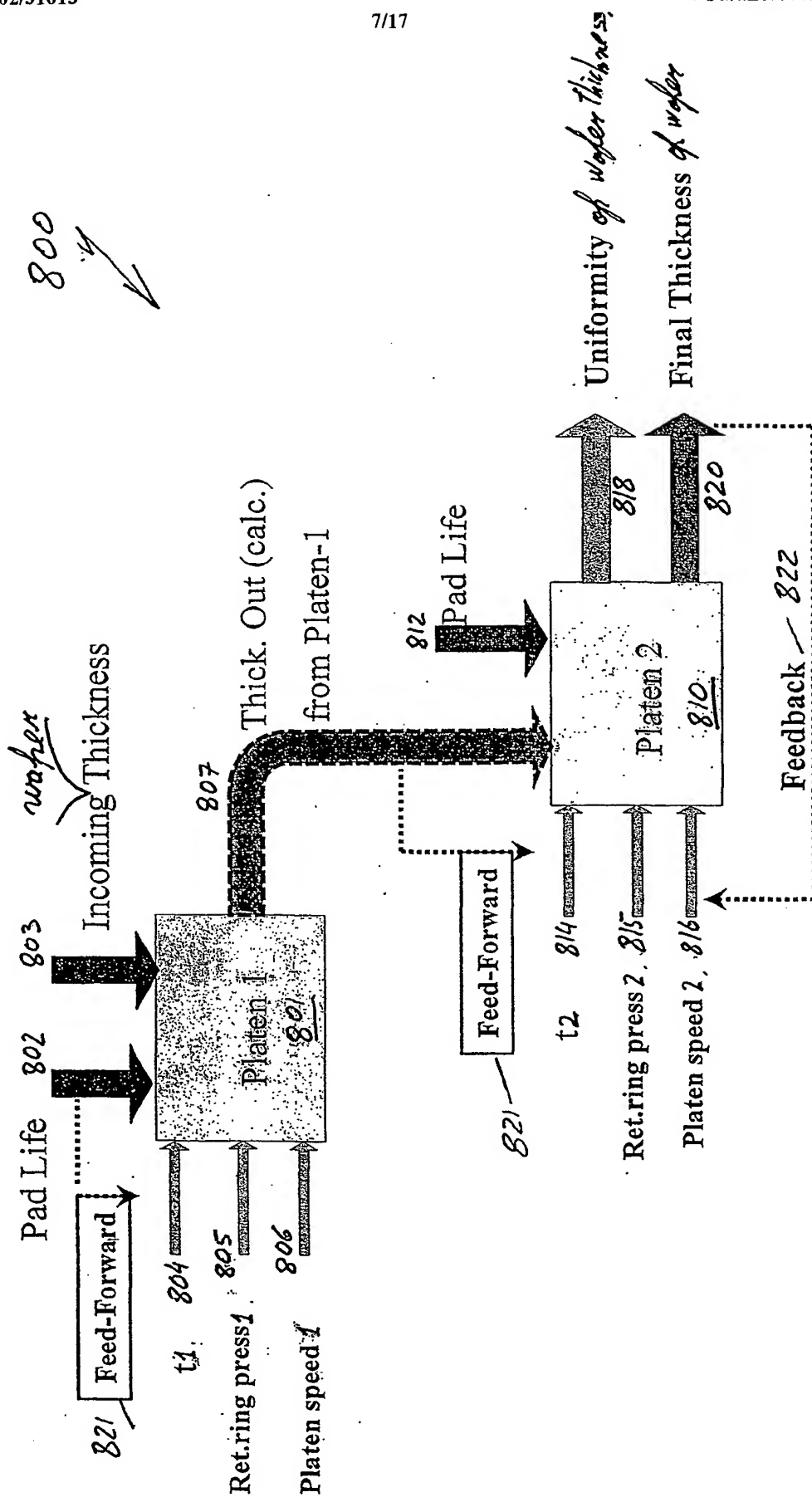


Figure 8



# Model Raw data

900  
9

919

918

904

902

906

905

112 Rows	5 Cols	RR press	Platen speed	Pad life (sec)	Polish Time	Uniformity	Avg. Removal
1	1	6.4	83.0	3171	64	7.1	3272.0
2	2	6.9	103.0	3779	60	9.5	3469.0
3	3	6.5	83.0	11116	55	5.3	2776.0
4	4	7.2	83.0	3307	64	14.9	3430.0
5	5	6.8	103.0	3714	60	6.5	3266.0
6	6	7.7	103.0	3714	60	9.8	3532.0
7	7	6.5	123.0	3141	64	20.4	4022.0
8	8	6.5	83.0	11057	55	3.4	2604.0
9	9	7.3	123.0	3013	55	7.6	3686.0
10	10	7.3	103.0	3563	60	7.8	3523.0
11	11	7.3	123.0	3345	64	25.6	4349.0
12	12	6.5	123.0	2954	55	8.1	3401.0
13	13	6.5	123.0	11154	55	5.4	3375.0
14	14	6.5	123.0	2895	55	12.2	3397.0
15	15	6.4	83.0	2858	55	3.0	2656.0
16	16	6.5	123.0	3209	64	20.1	4120.0
17	17	7.3	83.0	11175	55	3.5	2892.0
18	18	6.9	146.9	3843	60	12.8	4120.0
19	19	7.2	83.0	3239	64	7.0	3333.0
20	20	6.5	123.0	11095	55	9.7	3371.0
21	21	6.8	53.0	3778	60	5.9	2329.0
22	22	7.2	83.0	3035	55	8.9	2968.0
23	23	6.4	83.0	3103	64	6.4	2995.0
24	24	7.3	123.0	3277	64	29.1	3900.0
25	25	7.3	123.0	3072	55	8.1	3690.0
26	26	6.1	103.0	3660	60	11.5	3282.0
27	27	6.4	83.0	2917	55	4.8	2835.0
28	28	7.2	83.0	2978	55	3.0	2759.0
29	29	7.2	103.0	3588	45	3.7	2620.0
30	30	7.3	83.0	11234	55	8.0	2932.0
31	31	7.3	123.0	11272	55	13.4	3671.0
32	32	6.5	83.0	11302	64	4.3	3104.0
33	33	6.5	123.0	11340	64	10.4	3928.0

920

Figure 9

1000  
↙

# POEM model

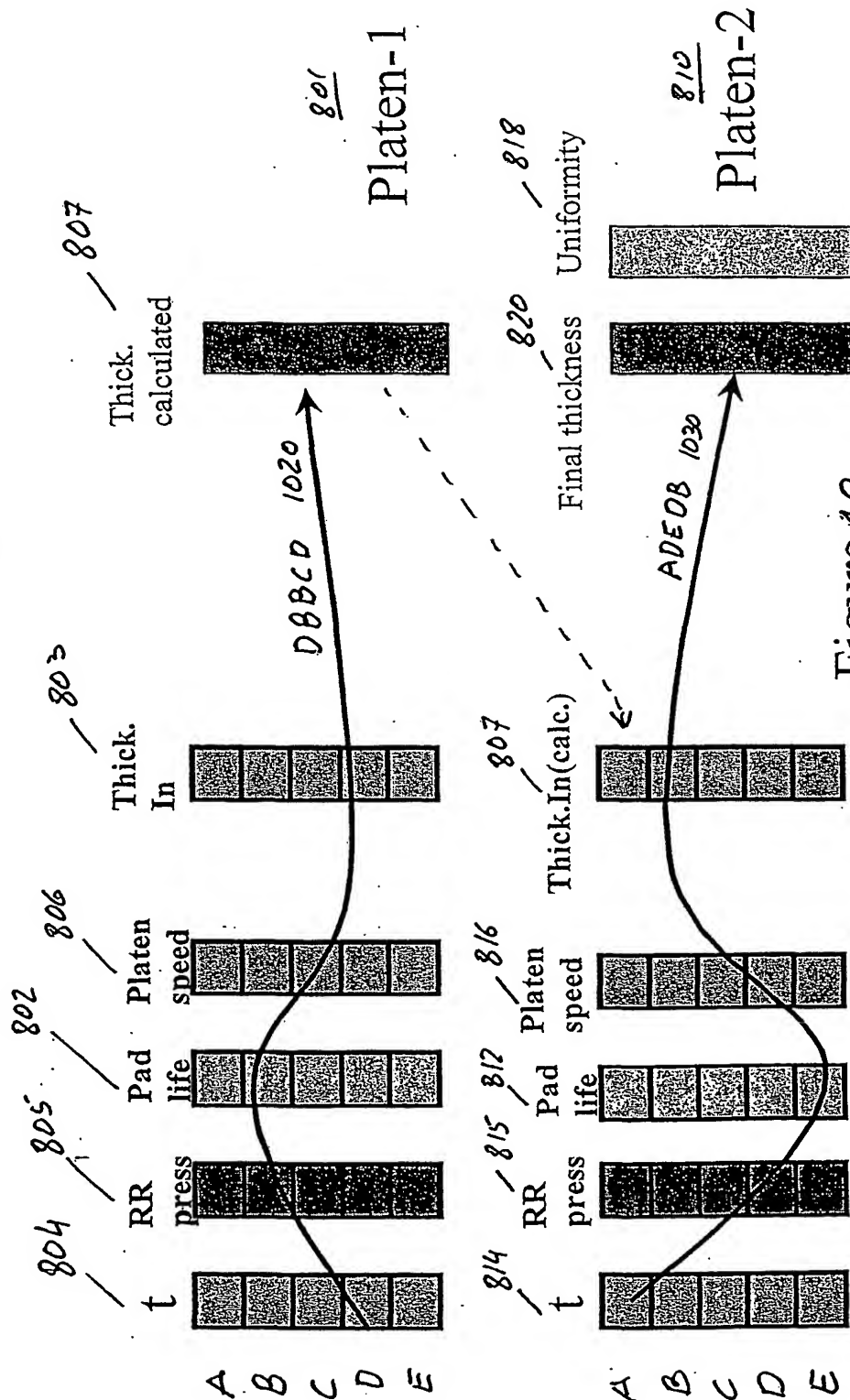


Figure 10

# POEM- Vectors Lookup Table

input	Vector value	n	sigma	Sigma std_error	Mean std_error	Vector code
83	2700.0	241	250	62	52	BBBBB
83	3300.0	326	201	65	55	FFFFF
83	3948.9	250	235	63	53	BBBBB
83	4148.9	300	237	71	61	BBBBD
83	4348.9	200	195	52	42	BBBBE
83	4548.9	261	215	62	52	BBBBF
83	3702.9	270	201	64	54	BBBBB
83	3902.9	301	200	71	61	BBBCC
83	4102.9	205	197	59	49	BBBCD
83	4302.9	268	231	63	53	BBBCE
83	4502.9	266	182	65	55	BBBDF
83	3656.9	310	201	70	60	BBBDB
83	3856.9	295	221	69	59	BBBDC
83	4056.9	195	301	72	62	BBBDD

Vector code in discretization symbols

Vector sigma

# of vector in population

Actual vector value

Input Name

Figure 11

# Model-Process Window, 1

BEST AVAILABLE COPY

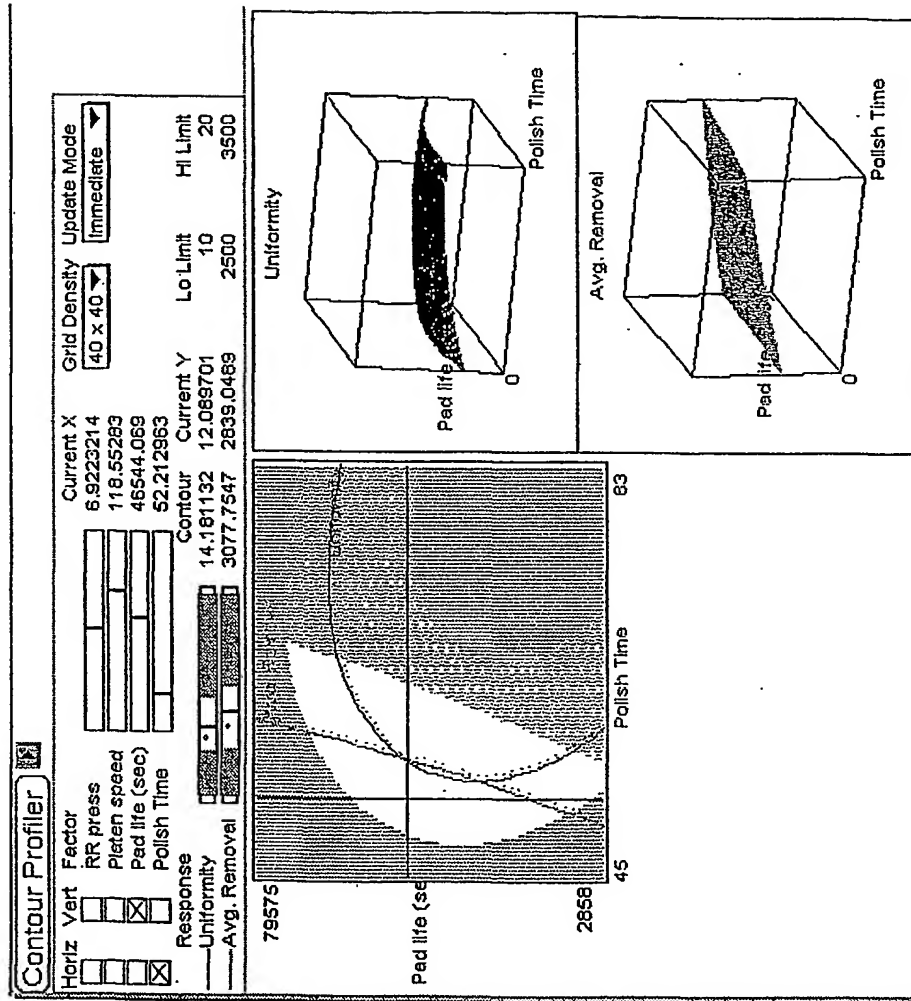
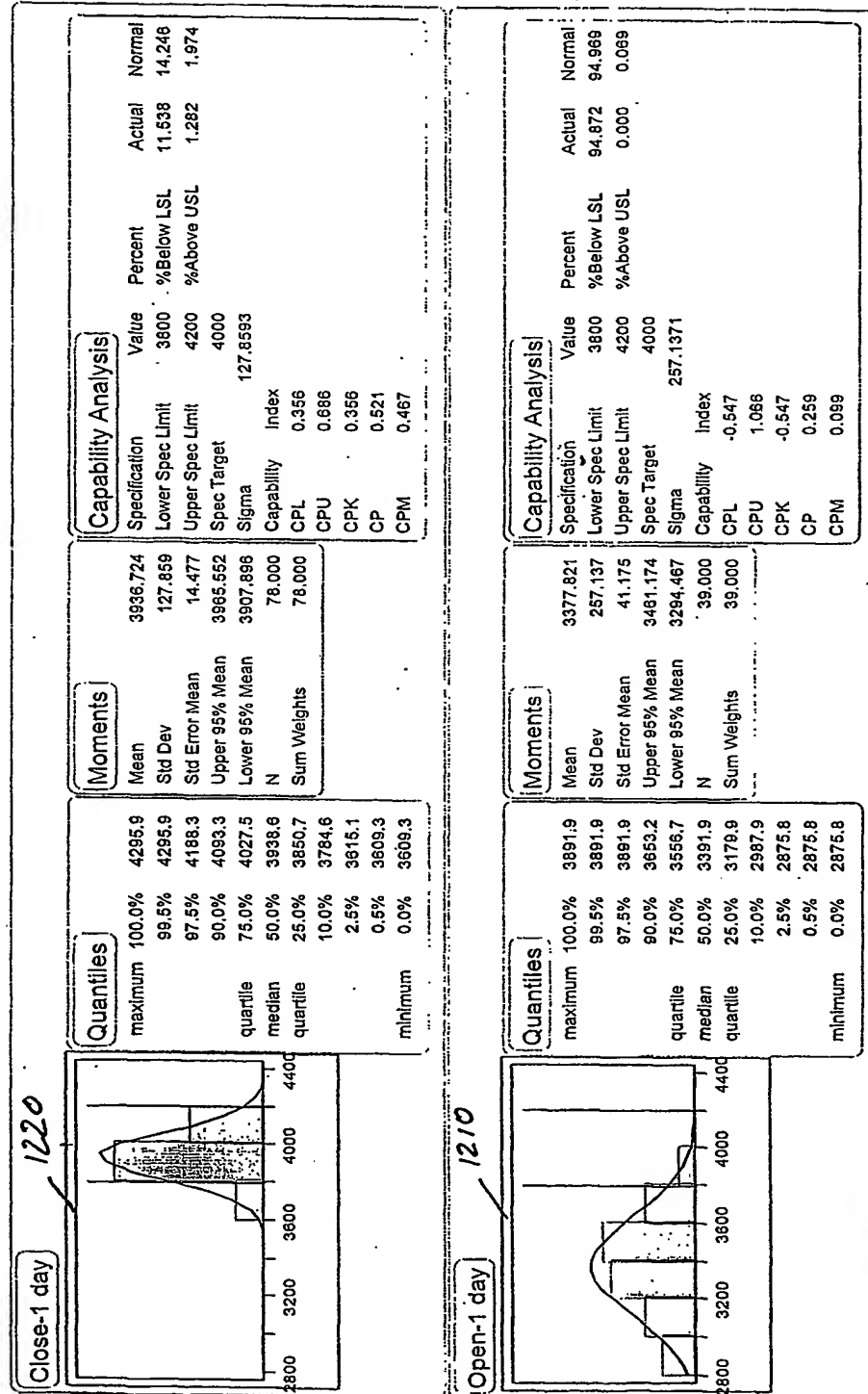


Figure 12

12/17

# Open vs Close loop Using POEM

1200



50 %

Improvement  
vs Open Loop

In sigma  
terms

Figure 13

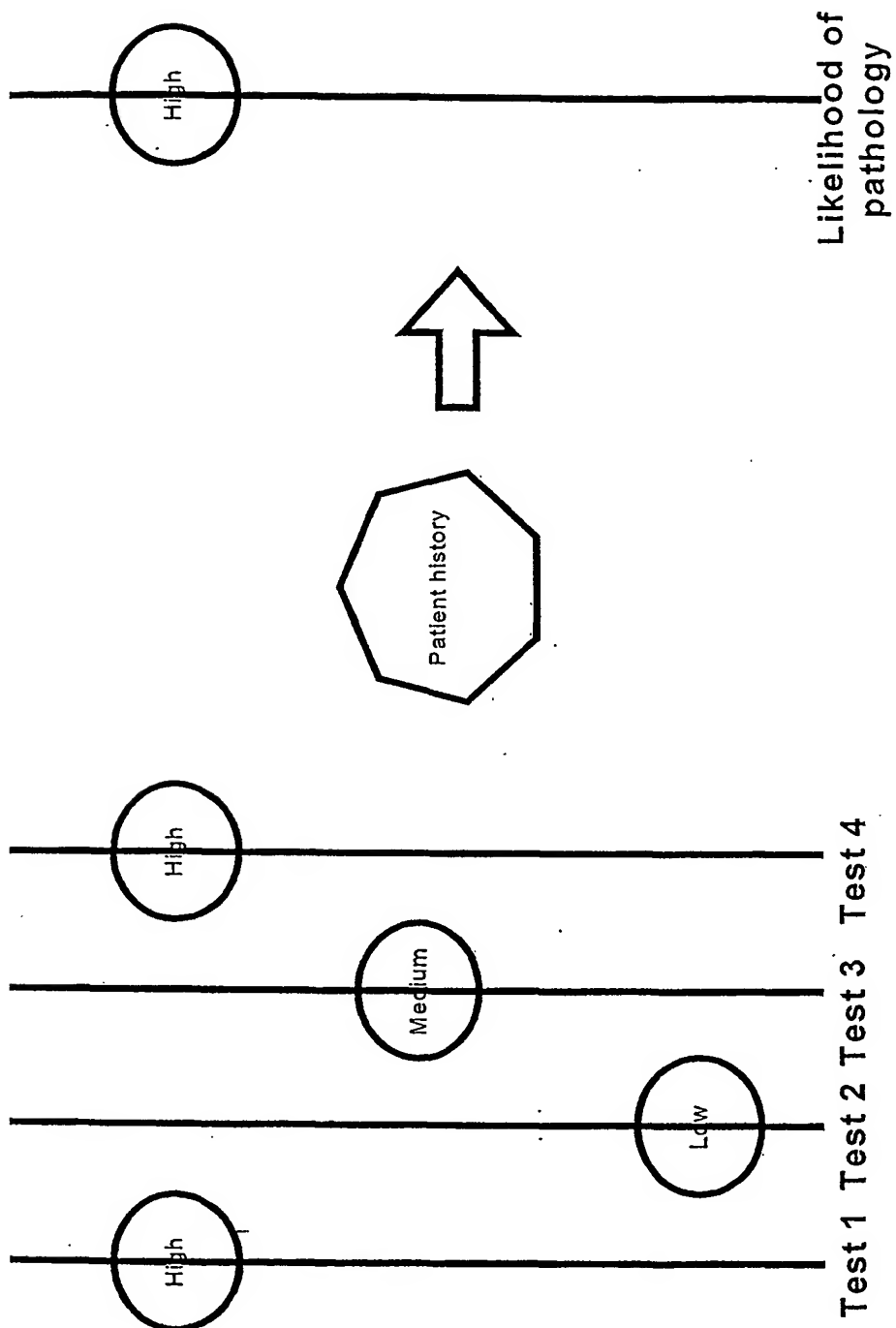


Figure 14

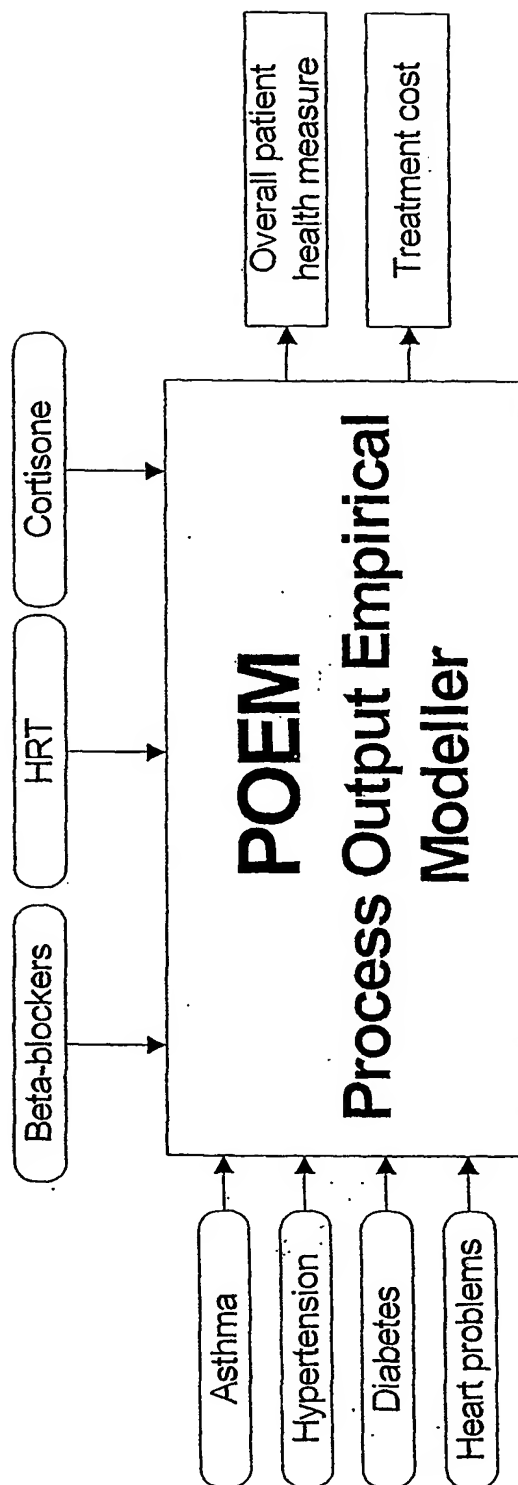


Figure 45

Figure 10a

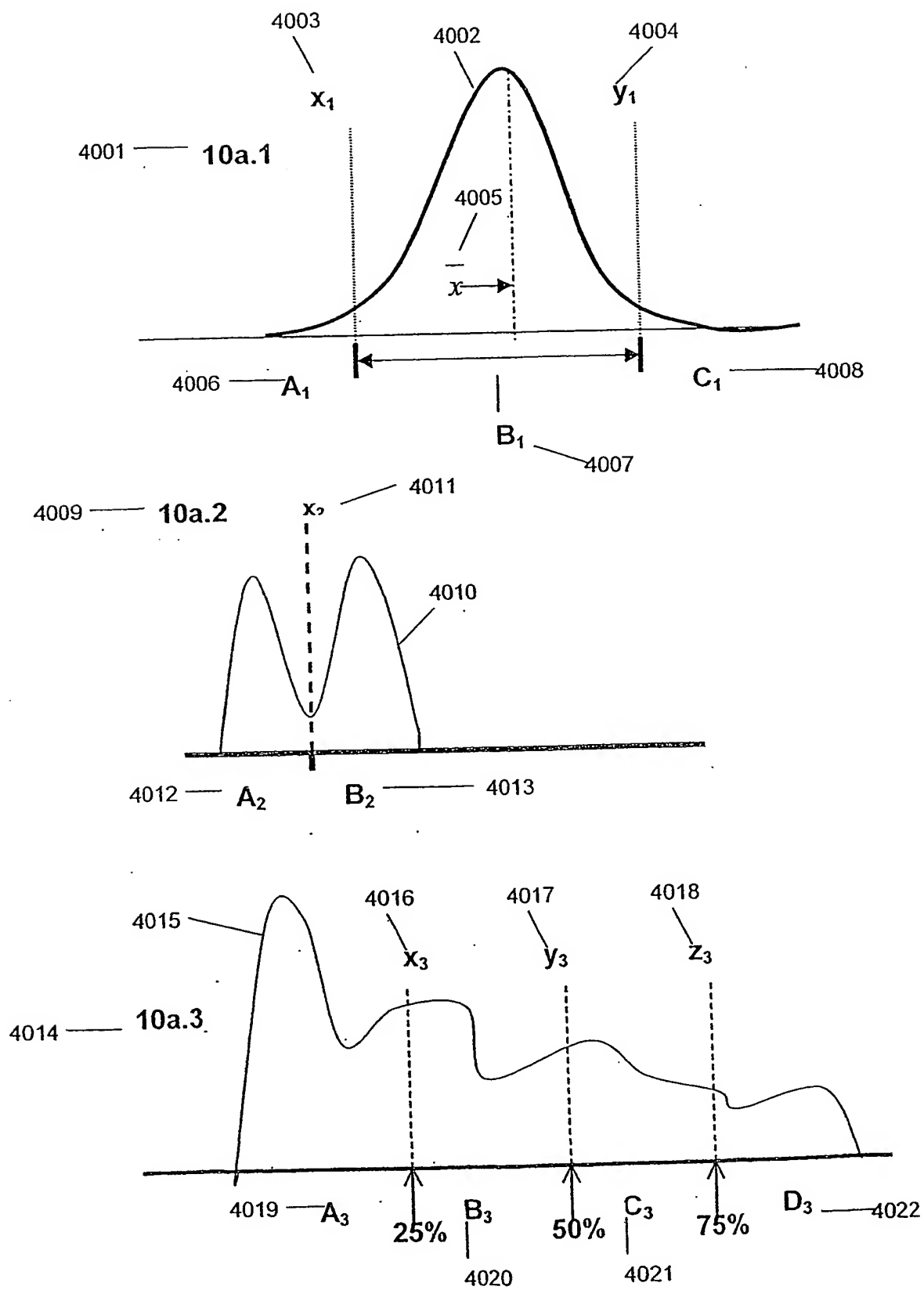




Figure 16b

5001	5002	5006
Process Runs	Input	Output
	Constant	Variables
	10a.1	10a.2 10a.3
1	5003	5004 5005
2	B <sub>1</sub>	A <sub>2</sub> D <sub>3</sub> O <sub>2</sub> 5011
3		
⋮		
e	B <sub>1</sub>	A <sub>2</sub> D <sub>3</sub> O <sub>e</sub> 5012
e+1	B <sub>1</sub>	A <sub>2</sub> D <sub>3</sub> O <sub>e+1</sub> 5013
⋮		
f	B <sub>1</sub>	A <sub>2</sub> D <sub>3</sub> O <sub>f</sub> 5014
⋮		

5000

5001

5002

5006

5007

5008

5009

5010

O<sub>2</sub>

O<sub>e</sub>

O<sub>e+1</sub>

O<sub>f</sub>

